PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-175787
(43)Date of publication of application : 24.06.1994

(51)Int.CI. G06F 3/06
G06F 12/00

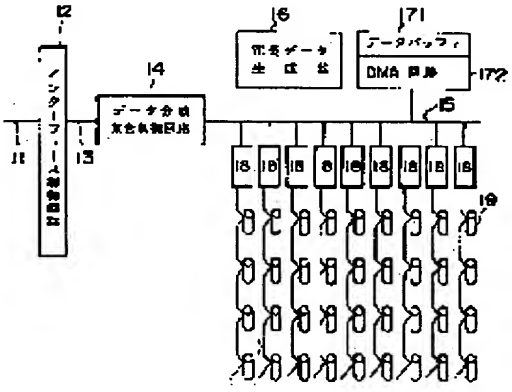
(21)Application number : 04-330739
(22)Date of filing : 10.12.1992

(71)Applicant : HITACHI LTD
HITACHI COMPUTER PERIPHERALS CO LTD

(72)Inventor : IWASAKI HIDEHIKO
TOKIDA YOSHINORI
MIZUNO KATSUTOSHI
SUZUKI RYOICHI
BABA HIDEMI

(54) DISK ARRAY DEVICE

(57)Abstract:
PURPOSE: To shorten the transfer time of redundant data to a disk device by generating redundant data in a unit smaller than a data distribution unit.
CONSTITUTION: Data inputted to a data distributing cluster control circuit 14 is divided into 512 bytes as a minimum distribution unit and is distributed in the 2n-fold distribution unit. A redundant data generator 16 manages distributed data in 512-byte unit by a higher-order data management counter, and generated data is managed in 512-byte units also by a redundant data management counter, and an input/output control circuit controls permission/inhibition of data input to the redundant data generator 16 while referring to both of them. An input/ output permission/inhibition signal as the result is transmitted to a DMA circuit 172 in each 512-byte unit. The DMA circuit 172 refers to transmitted information to determine the service of data transfer from a disk control circuit.



LEGAL STATUS

[Date of request for examination]	13.01.1999
[Date of sending the examiner's decision of rejection]	26.09.2000
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3186272
[Date of registration]	11.05.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開平6－175787
(43)公開日　平成６年(1994)６月24日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	３０１　Ｚ	7165－5B		
12/00	５４５　Ａ	8526－5B		

審査請求　未請求　請求項の数８(全 11 頁)

(21)出願番号　特願平4－330739

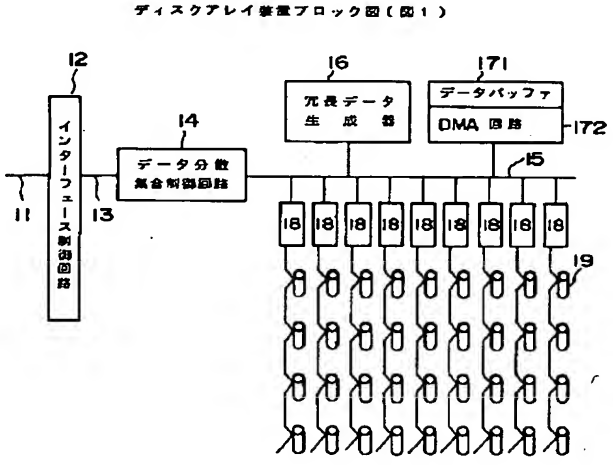
(22)出願日　平成４年(1992)12月10日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(71)出願人　000233033
日立コンピュータ機器株式会社
神奈川県小田原市国府津2880番地
(72)発明者　岩崎　秀彦
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(72)発明者　常田　義則
神奈川県小田原市国府津2880番地　日立コンピュータ機器株式会社内
(74)代理人　弁理士　富田　和子

最終頁に続く

(54)【発明の名称】　ディスクアレイ装置

(57)【要約】
【目的】　ディスクアレイ装置において、冗長データ生成器へのデータ入力、及び、冗長データ生成器からのデータ出力の転送効率を向上する。
【構成】　冗長データ生成器１６は、データ分散、集合制御回路１４、及び、ディスク制御回路１８からのデータを、分散単位よりも小さい単位にて入出力の可／不可制御をそれぞれ独立に行い、その制御結果をＤＭＡ回路１７２に伝達する。ＤＭＡ回路１７２は、これを元に、データ分散集合制御回路１４、及び、ディスク制御回路１８に対するデータ転送のサービスを行う。



【特許請求の範囲】

【請求項１】複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、
上記冗長データ生成手段は、冗長データを生成する際、データ分散単位よりも小さい単位で、冗長データを生成することを特徴とするディスクアレイ装置。

【請求項２】複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、
上位装置、及びディスク装置ユニットの両方から冗長データ生成手段にデータを転送して、冗長データを生成する場合に、上位装置から転送されてくるデータ量をカウントする上位側データ管理手段と、
ディスク装置ユニットから転送されてくるデータ量をカウントするディスク側データ管理手段と、
冗長データ生成手段が生成した冗長データ量をカウントする冗長データ管理手段と、
上記上位側データ管理手段と上記ディスク側データ管理手段と上記冗長データ管理手段とがカウントした結果に従って、上記冗長データ生成手段へのデータ転送を制御する入出力制御手段とを有することを特徴とするディスクアレイ装置。

【請求項３】複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、
データを上位装置に転送する時に、冗長データ及び冗長データを生成するために使用した上記データにより、誤りの有無を調べる誤り検出手段を有することを特徴とするディスクアレイ装置。

【請求項４】複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、
データをディスク装置ユニットに書き込む際に、書き込むデータの量が分散単位よりも小さいときに、
上記ディスク装置ユニットは、分散単位中の書替の対象となる記憶位置にあるデータはそのまま転送し、分散単位中の書替の対象とならない記憶位置にあるデータは転送せず、
上記冗長データ生成手段は、上記ディスク装置ユニットから転送されてきたデータ及び書き込むデータから冗長データを生成することを特徴とするディスクアレイ装置。

【請求項５】複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、
上記データを記録する時に、上記データに記録される位置を示す分散パラメータを付加する分散パラメータ付加手段を有し、
上記ディスク装置ユニットは、上記データとともに分散パラメータを記録することを特徴とするディスクアレイ装置。

【請求項６】請求項５記載のディスクアレイ装置において、
上記データを再生する時に、データと共に分散パラメータを各々のディスク装置ユニットより読みだし、分散パラメータが読みだした位置と整合が取れているがどうかを確認する分散パラメータ確認手段を有することを特徴とするディスクアレイ装置。

【請求項７】請求項５または６記載のディスクアレイ装置において、
上記冗長データ生成手段は、上記データより冗長データを生成する際、分散パラメータに対しても冗長データを生成し、分散パラメータを記録したディスク装置ユニットに障害が発生した時は、正常なディスク装置ユニットのデータと上記冗長データより、分散パラメータを再生成することを特徴とするディスクアレイ装置。

【請求項８】請求項５、６または７記載のディスクアレイ装置において、
上記データの記録される位置を示す位置情報を保持する位置保持手段と、
上記位置保持手段が保持する位置情報が消失した場合、ディスク装置ユニットに記録された分散パラメータにより位置情報を生成する位置情報生成手段とを有することを特徴としたディスクアレイ装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、複数台のディスク装置ユニットにデータを分散し、その分散されたデータに対して、冗長データを生成するディスクアレイ装置に関する。

【０００２】

【従来の技術】ディスクアレイ装置に関する従来技術としては、例えば、特開平２－２３６７１４号公報「アレイ型ディスク駆動機構システム及び方法」等で示されるように、データを複数台のディスク装置ユニットに分散して記憶し、高速転送を実現するものや、分散データより冗長データを生成し、ディスク障害時、冗長データよりデータを再構築するものがある。

【０００３】

【発明が解決しようとする課題】冗長データを有するディスクアレイ装置では、データライト時の冗長データの生成、及び生成した冗長データのディスク装置への転送が必要となり、このことがデータ転送効率を低下させる一因となる。

【０００４】本発明の第１の目的は、冗長データのディスク装置への転送時間を短縮したディスクアレイ装置を提供することである。

【０００５】また、データライト時に生成した冗長データは、データリード時には、リードされず、ディスク障害発生時のデータ再構築時にリードされる。このため、データリードを主に行われるディスクアレイ装置においては、冗長データ用ディスクへのアクセスが少ないため、障害検出が遅れ、２台のディスク装置ユニットが障害を起こす場合がある。２台のディスク装置ユニットが障害を起こすと原理的にデータ再構築不能となる。

【０００６】本発明の第２の目的は、上記問題点を解決するため、障害検出の遅れを防ぎ、信頼性を向上させたディスクアレイ装置を提供することにある。

【０００７】また、ディスクアレイ装置におけるデータは、冗長データを有することで、ディスク障害時等の信頼性を確保しているが、データの分散および集合はディスクアレイ装置内の回路が実行する為、分散、集合させる回路そのものに障害があった場合、または、分散記録した時のパラメータ（分散させたディスク装置ユニットの番号等）が、分散、集合させる回路内より消失した場合、データ障害の発生がなくともデータ破壊となる。

【０００８】本発明の第３の目的は、分散、集合に対する信頼性を向上させたディスクアレイ装置を提供することにある。

【０００９】

【課題を解決するための手段】上記第１の目的を達成するために、複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、上記冗長データ生成手段は、冗長データを生成する際、データ分散単位よりも小さい単位で、冗長データを生成することとした。

【００１０】上記第２の目的を達成するために、複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、データを上位装置に転送する時に、冗長データ及び冗長データを生成するために使用した上記データにより、誤りの有無を調べる誤り検出手段を有することとした。

【００１１】上記第３の目的を達成するために、複数のディスク装置ユニットを有し、各々のディスク装置ユニットに上位装置から転送されてくるデータをデータ分散単位で分散して、記録するとともに、上記データに対して冗長データを生成する冗長データ生成手段を有するディスクアレイ装置において、上記データを記録する時に、上記データに記録される位置を示す分散パラメータを付加する分散パラメータ付加手段を有し、上記ディスク装置ユニットは、上記データとともに分散パラメータを記録することとした。

【００１２】

【作用】ディスクアレイ装置において、データライト時、冗長データ生成手段により、冗長データを生成する際、データ分散単位よりも小さい単位で、冗長データを生成する。

【００１３】冗長データ生成手段において、上位装置からのデータはデータ分散単位を分割した単位で入力され、また、生成した冗長データは、同様にデータ分散単位を分割した単位で出力制御される。このため、冗長データは、分散単位分の冗長データが生成されるのを待つことなく、１つの分割単位の冗長データが生成される毎に出力し、また、上位装置からのデータは、データ分散単位＊分散ディスク装置数のデータ転送を行った後、生成した分散単位の冗長データが全て冗長データ生成手段より出力されるのを待つことなく、生成単位の冗長データが出力された時点で次の分散データの入力を行うことができる。

【００１４】また、ディスクアレイ装置において、誤り検出手段は、データを上位装置に転送する時に、冗長データ及び冗長データを生成するために使用した上記データにより、誤りの有無を調べる。

【００１５】このため、データリード時に冗長データも含めてリードし、データ照合するため、早い時点で冗長データが記録されたディスク装置ユニットの故障が検知でき、ディスク装置ユニットが２台故障することを防げる。従って、信頼性が向上する。

【００１６】さらに、ディスクアレイ装置において、分散パラメータ付加手段は、上記データを記録する時に、上記データに記録される位置を示す分散パラメータを付加し、上記ディスク装置ユニットは、上記データとともに分散パラメータを記録する。

【００１７】このため、データ記録時、分散パラメータが生成、付加され、その結果、分散先のディスク装置ユニット毎に、付加された分散パラメータを照合し、分散データの妥当性を保証して分散データと共にディスク装置に記録する等のことが可能となる。

【００１８】

【実施例】以下、添付の図面に示す実施例により、更に詳細に本発明について説明する。図１は、本発明を説明するための冗長データを有するディスクアレイ装置の第１の実施例である。また、図２は、冗長データ生成手段

および誤り検出手段である冗長データ生成器の入出力制御ブロック図である。

【００１９】本ディスクアレイ装置は、ホストＳＣＳＩバス１１と、インタフェース制御回路１２と、データバス１３と、データ分散集合制御回路１４と、共通データバス１５と、データバッファ１７１と、ＤＭＡ回路１７２と、冗長データ生成器１６と、ディスク制御回路１８と、ディスク装置ユニット（ディスク駆動機構）１９とを有する。冗長データ生成器１６は、上位側データ管理手段である上位側データ管理カウンタ２１と、ディスク側データ管理手段であるディスク側データ管理カウンタ２２と、冗長データ管理手段である冗長データ管理カウンタ２３と、入出力制御手段である入出力制御回路２４とを有する。

【００２０】本ディスクアレイ装置は、コンピュータの記憶装置として動作し、汎用Ｉ／Ｏインターフェースであるホスト ＳＣＳＩバス１１を介して、コンピュータと接続される。

【００２１】コンピュータのデータ書込み要求時、ホストＳＣＳＩバス１１を介して転送されるデータは、インターフェース制御回路１２を経由し、データバス１３よりデータ分散集合制御回路１４に入力され、ここで分散される。分散されたデータは、ＤＭＡ回路１７２により、共通データバス１５を介して、ディスク制御回路１８に解放されたデータバッファ１７１のエリアに転送される。この時、分散されたデータは、データバッファ１７１に転送すると同時に、冗長データ生成器１６にも転送される。転送された各々の分散データは、排他ＯＲが実行され、冗長データ生成器１６内に冗長データが生成される。冗長データの生成が完了すると、分散データと同様にして該当するデータバッファ１７１のエリアに転送される。各々のディスク制御回路１８は、自身に解放されたデータバッファ１７１のエリアより、指定されたディスク装置ユニット１９に対して、分散されたデータ、及び、分散データより生成された冗長データを書込む。

【００２２】コンピュータのデータ読みだし要求時は、該当するディスク制御回路１８が、指定されたディスク装置ユニット１９よりデータを読みだし、データバッファ１７１の定められたエリアに転送すると同時に、冗長データ生成器１６にも転送する。冗長データ用のディスク制御回路１８は、冗長データ生成器１６にのみ転送する。データ分散集合制御回路１４は、データバッファ１７１に転送された分散データを集合し、集合されたデータは、データバス１３、インターフェース制御回路１２を経由し、ホストＳＣＳＩバス１１を介しコンピュータに転送される。また、冗長データ生成器１６は、入力された分散データ、及び冗長データの排他ＯＲを実行し、その結果がオール“００”であることを確認し、分散データと冗長データの整合性をチェックする。

【００２３】分散データから生成され、ディスク装置ユニット１９に書込まれた冗長データは、分散を行った複数のディスク装置ユニット１９の中で、あるディスク装置ユニット１９に障害が発生した場合の、障害が発生したディスク装置ユニット１９の分散データを復元するために使用される。分散データの復元は、障害の発生したディスク装置ユニット１９以外のディスク装置ユニット１９より、分散されたデータ、及び冗長データを読みだし、冗長データ生成器１６へ転送（分散データは、該当するデータバッファエリア１９にも転送）することで行う。

【００２４】冗長データ生成器１６は、転送されたデータの排他ＯＲを実行し、障害分散データを再構築する。再構築されたデータは、該当するデータバッファ１７１のエリアに転送され、データ分散集合制御回路１４は、通常の読みだし処理と同様に、データバッファ１７１の分散データを集合し、集合されたデータは、コンピュータへと転送される。また、ディスクアレイ装置内にスペアディスク装置ユニットを備えている場合は、ディスクアレイ装置内で、障害ディスク装置ユニットの内容をスペアディスク装置ユニットに再構築する。

【００２５】コンピュータのデータ書込み要求時で、各々のディスク装置ユニット列の途中の列からデータ書込みが開始され、なおかつ、その分散単位の途中からデータ書込みが開始される場合について、更に詳細に説明する。この場合、コンピュータから転送されるデータからのみでは冗長データは生成できないため、ディスク装置ユニット１９からのデータ読みだしも行わなければならない。

【００２６】ホストＳＣＳＩバス１１、インターフェース制御回路１２、データバス１３を介し、データ分散集合制御回路１４に入力されたデータは、ここで最小分散単位である５１２バイトに分割され、なおかつ、その２ｎ 倍の分散単位にて分散される。

【００２７】また、この時分散単位の途中からデータ書込み開始となるため、データ分散集合制御回路１４は、分散単位の開始点から書込み開始までを“００”データを付加し転送する。分散されたデータは、ＤＭＡ回路１７２によりデータバッファ１７２、及び、冗長データ生成器１６に、５１２バイト単位で転送される。但し、付加された“００”データは、冗長データ生成器１６にのみ転送される。この時、冗長データ生成器１６は、上位側データ管理カウンタ２１により、分散データを５１２バイトの単位に管理し、また、生成した冗長データについても、冗長データ管理カウンタ２３により、５１２バイト単位に管理し、双方を参照しながら、入出力制御回路２４は、冗長データ生成器１６へのデータ入力の可／不可制御を行う。

【００２８】その結果である入出力可／不可信号２５は、ＤＭＡ回路１７２に対し、５１２バイト単位毎に伝

達される。

【００２９】ＤＭＡ回路１７２は、その伝達された情報を参照し、データ分散集合制御回路１４からのデータ転送のサービスを決定する。一方、該当するディスク制御回路１８は、指定されたディスク装置ユニット１９よりデータを読みだし、５１２バイト単位に分割する。

【００３０】また、この時、分散単位の途中からデータ書込み開始となるディスク制御回路１８は、分散単位の開始点から書込み開始点までを、“００”データを付加し転送する。分割されたデータは、ＤＭＡ回路１７２により、冗長データ生成器１６にを転送されるこの時、冗長データ生成器１６は、ディスク側データ管理カウンタ２２により、ディスク制御回路１８からのデータを、ディスク制御回路１８毎に５１２バイトの単位に管理し、また、生成した冗長データについても、冗長データ管理カウンタ２３により、５１２バイト単位に管理し、双方を参照しながら、冗長データ生成器１６へのデータ入力の可／不可制御を行う。

【００３１】また、上位側データ管理カウンタ２１により、データ分散集合制御回路１４からのデータも合わせて管理しており、冗長データ生成器１６は、入力されるデータ全てを独立に制御するため、データ転送にソフト制御を必要としない。

【００３２】冗長データ生成器１６へのデータ入力の可／不可制御の結果は、入出力可／不可信号２５として、ＤＭＡ回路１７２に対し、５１２バイト単位毎に伝達される。ＤＭＡ回路１７２は、その伝達された情報を参照し、ディスク制御回路１８からのデータ転送のサービスを決定する。

【００３３】データ分散集合制御回路１４、及び、ディスク制御回路１８から冗長データ生成器１６に入力されたデータは、配置ＯＲが実行され、冗長データ生成器１６内に、新しい冗長データが生成される。この時、分散データが欠けている部分についても、“００”のダミーデータが転送されているため、排他ＯＲを行う列数は、分散データが欠けている部分、揃っている部分共に同一列数として扱う。また、分散データが欠けている部分については、“００”データにて排他ＯＲを実行するため、この部分については、排他ＯＲを実行しないことと同様であり、ダミーデータを含んでいても、冗長データは破壊されることはない。冗長データの生成が完了すると、該当するデータバッファ１７１のエリアに転送される。各々のディスク制御回路１８は、自身に解放されたデータバッファ１７１のエリアより、指定されたディスク装置ユニット１９に対して、分散されたデータ、及び、新しく生成された冗長データを書込む。

【００３４】以上のように、本発明は冗長データを有するディスクアレイ装置において、冗長データ生成器へのデータ入力、及び、冗長データ生成器からのデータ出力の、転送効率の向上を可能とする。

【００３５】また、データリード時、冗長データも含めてリードするため、冗長データの信頼性を向上させることが可能となる。

【００３６】本発明によれば、データライト時、冗長データ生成器により、上位装置からのデータはデータ分散単位を分割した単位で入力可／不可制御され、また、生成した冗長データは、同様にデータ分散単位を分割した単位で出力制御される。このため、冗長データは、分散単位分の冗長データが生成されるのを待つことなく、１つの分割単位の冗長データが生成される毎に出力し、また、上位装置からのデータは、データ分散単位＊分散ディスク装置数のデータ転送を行った後、生成した分散単位の冗長データが全て冗長データ生成器より出力されるのを待つことなく、分割単位での冗長データ出力に応じた次の分散のデータ入力を行うことができる。

【００３７】すなわち、冗長データ生成器へのデータの入出力可／不可制御を、データ分散単位を分割した単位で行うことで、生成した冗長データの出力待ちによるデータ入力不可時間を低減し、データ転送の効率を向上させる。

【００３８】また、各々のディスク装置ユニット列の途中の列にてデータライトが開始、あるいは、終了するような場合、この分散データに対する冗長データを生成する場合は、上位装置、及びディスク装置からの冗長データ生成器へのデータ入力が必要であるが、この時、上位装置、各ディスク装置からのデータ入力順、及び、タイミングに制限がないため、データ入力にソフト介在の必要がなく、また、上位装置とディスク装置と同時に転送起動可能となり、ディスク装置からのデータ入力のオーバヘッドが低減できる。

【００３９】すなわち、冗長データ生成器へのデータの入出力可／不可制御を、上位装置側、及び、それぞれのディスク装置側全て独立に制御することで、冗長データ生成器へのデータ入力タイミング、データ入力順（上位装置側及びディスク装置側のどちらを先に入力させるかという順序）等を意識する必要がなくなり、またそのためのソフト制御も必要とせず、データ転送の効率を向上させる。

【００４０】更に、分散単位の途中にてデータライトが開始、あるいは終了するような場合に、分散単位の開始から途中まで、あるいは、分散単位の途中から終了までをダミー転送することで、上位装置、及び、ディスク装置共にデータ転送の単位は分散単位となり、冗長データ生成のための転送を、分散単位の開始から途中までの列、あるいは、分散単位の途中から終了までの列、及び、それ以外の列とに分割して制御する必要がなくなる。

【００４１】すなわち、従来ホストマシンから送られてくるデータの単位よりも分散単位が大きいときに分散単位分のデータをディスク装置ユニットから実際に読み込

まなければならなかったが、ダミーデータを使用することにより、実際に読み込まなければならないデータを減らすことができる。従って、冗長データの生成効率が向上する。

【００４２】また、データリード時に冗長データも含めてリードし、冗長データ生成器にてデータ照合するため、データ再構築時の信頼性を向上できる。

【００４３】次に、第２の実施例について説明する。

【００４４】図３は、本発明を説明する為の、ディスクアレイ装置の実施例である。本実施例の装置は、コンピュータの記憶装置として動作する。本実施例では、汎用Ｉ／Ｏインターフェースであるホスト ＳＣＳＩバス１１を介して、コンピュータと接続される。本装置の機能は、コンピュータの要求に対して、該当するディスク装置ユニットに、データ書き込み及び、読みだしを行なうことであるが、本機能を実現する手段に、データの分散集合手法を用いていることが第１の特徴としてあげられる。

【００４５】本ディスクアレイ装置は、ホストＳＣＳＩバス１１と、インタフェース制御回路１２と、データバス１３と、データ分散集合制御回路１１４と、共通データバス１５と、データバッファ１７１と、ＤＭＡ回路１７２と、冗長データ生成及びデータ再構築回路１６１と、ディスク制御回路１８と、ディスク装置ユニット（ディスク駆動機構）１９とを有する。

【００４６】本実施例においては、コンピュータのデータ書き込み要求時、ホストＳＣＳＩバス１１を介して転送されるデータは、インターフェース制御回路１２を経由し、データバス１３よりデータ分散集合制御回路１４に入力される。

【００４７】ここでコンピュータの指示した論理ユニットアドレス、論理ブロックアドレス、転送ブロック数より、どのディスク装置ユニットのどの位置に、どれだけのデータ転送を行なうか等の、分散パラメータが生成され、コンピュータからのデータが分散される。

【００４８】分散されたデータは、共通データバス１５を介して、分散先のディスク装置ユニット１９を制御するディスク制御回路１８に開放されたデータバッファ１７１のエリアにＤＭＡ回路１７２を介して転送される。各々のディスク制御回路は、自身に開放されたデータバッファのエリアより、指定されたディスク装置ユニット１９に対して、分散されたデータを書き込む。

【００４９】コンピュータの、データ読みだし要求時は、コンピュータの指示するパラメータ及び分散情報より、該当するディスク制御回路１８１が、指定されたディスク装置ユニット１９よりデータを読みだし、データバッファ１７１の定められたエリアに転送する。データ分散集合制御回路１４は、データバッファ１７１に転送された分散データを集合し、集合されたデータは、データバス１３、インターフェス制御回路を経由して、ホストＳＣＳＩバス１１を介してコンピュータに転送される。

【００５０】本ディスクアレイ装置の第２の特徴として、データの冗長性があげられる。ディスクアレイ装置に冗長性を持たす場合は、コンピュータのデータ書き込み要求時、データ分散集合制御回路１４により分散されるデータを、データバッファ１７１に転送すると同時に、冗長データ生成及びデータ再構築回路１６に転送する。転送された各々の分散データは、分散するディスク装置ユニット数分、排他ＯＲが実行されて、生成回路内に冗長データが生成される。

【００５１】本冗長データの転送先は、冗長データ生成及びデータ再構築回路１６にデータが入力される際に、データ分散集合制御回路１４１より、冗長データ生成及びデータ再構築回路１６に転送されているため、冗長データの生成が完了すると、分散データと同様にして該当するデータバッファ１７１のエリアに転送され、ディスク制御回路１８１を経由してディスク装置ユニットに書き込まれる。

【００５２】このように、分散データから生成され、ディスク装置ユニットに書き込まれた冗長データは、分散を行なった複数のディスク装置ユニットの中で、あるディスク装置ユニットに障害が発生した場合、障害の発生したディスク装置ユニットの分散データを復元するために使用される。

【００５３】本装置では、障害の発生したディスク装置ユニット以外のディスク装置ユニットより、分散されたデータ及び、冗長データを読みだし、データは、該当するデータバッファ１７１内のエリアと、冗長データ生成及びデータ再構築回路１６へ転送し、冗長データは、冗長データ生成及びデータ再構築回路１６へのみ転送することで、冗長データ生成及びデータ再構築回路１６は、障害分の分散データを再構築する。

【００５４】この再構築されたデータは、該当するデータバッファ１７１のエリアに転送される。データ分散集合制御回路１４１は、通常の読みだし処理と同様に、データバッファ１７１の分散データを集合し、集合されたデータは、コンピュータへと転送される。

【００５５】また、ディスクアレイ装置内に、スペアディスク装置ユニットを備えている場合は、ディスクアレイ装置内で、障害ディスク装置ユニットの内容を、スペアディスクに再構築する。

【００５６】以上、ディスクアレイ装置の実施例を説明したが、本発明は、コンピュータのデータを、データ分散、集合手段を用いて装置内で加工する為のデータ保証方法であり、その実施例を図４、図５に示す。

【００５７】図４は、データ分散集合制御回路１４１のブロック図を表わす。本データ分散集合制御回路１４１は、分散レングスカウンタ１２１と、分散パラメータ発生器（分散パラメータ付加手段）１２２と、バッファ分

散アドレス生成器１２３と、分散パラメータ照合器（分散パラメータ確認手段）１２４と、ＦＩＦＯメモリ１２３とを有する。

【００５８】データバス１３より、入力されたコンピュータからのデータは、分散レングスカウンター１２１により、最小分散単位の５１２バイトのデータ長に分割され、ＦＩＦＯメモリ１２５に入力される。このとき、分散されたデータに、分散パラメータ発生器１２２によって生成された分散パラメータ、本実施例では、８バイト長が、最小分散データ単位に付加されてＦＩＦＯメモリ１２５に入力される。ＦＩＦＯメモリ１２５に入力された各々の分散データ及び分散パラメータは、共通バス１５を介して、データバッファ１７１及び、冗長データ生成及びデータ再構築回路１１６へＤＭＡ転送される。このときのデータバッファ１７１転送アドレスは、バッファ分散アドレス生成器１２３によって生成される。

【００５９】コンピュータにデータを転送する際は、データバッファ１７１に転送される分散データを、バッファ分散アドレス生成器１２３によって生成されるバッファアドレスよりＦＩＦＯメモリ２５に転送する。ＦＩＦＯメモリ２５に転送された各々の分散データ及び、分散パラメータは、ＦＩＦＯメモリ２５上に分散された順序で集合しており、ＦＩＦＯメモリ２５よりデータバス１３にデータ転送をする際、分散パラメータ照合器２４が分散パラメータを照合して、分散データより切り離し、データのみをデータバス１３に転送する。

【００６０】図５は、ディスク制御回路１８１のブロック図を表わす。

【００６１】本ディスク制御回路１８１は、ＦＩＦＯメモリ３１と、分散パラメータ照合器３２と、ＳＣＳＩ制御器３３と、ディスクＳＣＳＩバス３４とを有する。

【００６２】共通データバス１５を介して入力される各々の分散データ及び、分散パラメータは、ＦＩＦＯメモリ３１に転送される。ＦＩＦＯメモリ３１に転送された分散データ及び分散パラメータは、分散パラメータ照合器３２で分散データの妥当性が照合され、ＳＣＳＩ制御器３３を介して、該当のディスク装置ユニットに、分散パラメータも含めて書き込まれる。

【００６３】ディスクＳＣＳＩバス３４より入力される分散データ及び分散パラメータは、ＳＣＳＩ制御器を介してＦＩＦＯメモリ３１へ転送される。このとき分散パラメータは、分散パラメータ照合器３２の妥当性が照合する。ＦＩＦＯメモリ３１に転送された分散データ及び分散パラメータは、共通データバスへ出力される。

【００６４】以上の様に、本発明は、ディスクアレイ装置において、データの分散、集合をおこなう際のデータの保証を可能とする。また、分散パラメータがディスク装置ユニットに書き込まれることから、分散パラメータを読み込むことで、分散情報を再生することが可能となる。

【００６５】本実施例における分散パラメータの内容を下記の表１に示す。

【００６６】

【表１】

表１　分散パラメータ構成

付加パラメータ名	ビット長	内　容
ディスク制御番号	4	ディスク装置ユニットを制御するディスク制御回路の固有番号。
論理ユニット番号	6	上位より指定の論理ユニットアドレス。
シーケンス番号	6	１回で分散する分散単位内のブロックの順位。
パリティグループ番号	３２	同一論理ユニット内の（分散単位）×（分散ディスク装置）分のデータを管理する番号。
ＬＲＣコード	１６	上記パラメータを含めた分散ブロック単位の冗長コード。

【００６７】なお、本実施例において、分散パラメータに規則性を持たせる、すなわち規則的にデータを分散させると、分散パラメータの排他論理和をとったものも予測可能なものとなる。これを冗長データの分散パラメータとみなして、冗長データに付加して記録し、再生時にこの分散パラメータが予測するものになっているかどう

かを判定することにより、冗長データを記録しているディスク装置ユニットに故障が生じたかどうかがわかる。従って、ディスクアレイ装置の信頼性が向上する。

【００６８】本実施例によれば、データ記録時、最小分散長である５１２バイトのデータ単位に分散パラメータが生成、付加され、分散先のディスク装置ユニット毎に、付加された分散パラメータが照合され、分散データの妥当性を保証して分散データと共にディスク装置に記録される。また冗長データにおいても、分散パラメータ分の冗長データが生成され、冗長データとしての妥当性を照合し、ディスク装置に記録される。データ読みだし時は、ディスク装置に記録されたデータと共に分散パラメータが読みだされ、集合時にデータの妥当性を保証して、正当なデータが上位システムに転送される。また、障害データを他のデータと冗長データより再構築する場合においても、障害データの分散パラメータが他のデータの分散パラメータと冗長データより再構築されるため、集合時にデータの妥当性を保証することが可能となる。

【００６９】さらに、分散パラメータを、ディスク装置上に記録するため、ディスクアレイ装置内に保持している分散情報が消失した場合においても、ディスク装置が正常であれば、ディスク装置に記録された分散パラメータより分散情報の再構築が可能となる。

【００７０】さらに、以下のような効果もある。

【００７１】通常、コンピュータから送られるデータは、ブロック単位（分散単位よりも小さい単位。例えば、ホストシステムからの転送単位が５１２バイトで、分散単位が２Ｋバイトのときの５１２バイト）にチェックコード等がディスクアレイ装置側により付加され、読みだし時に照合されるが、ディスクアレイ装置では、コンピュータからのデータを装置内で分散集合加工するため、ブロック単位のチェックコード等は正常でも、ブロック単位の分散単位内の順序（シーケンス）が故障により保証されなくなる。分散パラメータの中にシーケンス情報を含ませることにより、シーケンスもチェックすることができる。

【００７２】また、ディスクアレイ装置内で生成される冗長データにも分散パラメータを付加し、冗長データより、障害データを再構築する際、再構築したデータの分散パラメータをも再構築し、分散パラメータより故障したディスク装置ユニットに分散されたものであるかどうかがわかる。もし、故障したディスク装置ユニット以外に分散されたものであるという分散パラメータが検出されたときは、ユニットの故障以外が考えられることにより、再構築されたデータの信頼性が向上する。

【００７３】さらに、ディスクアレイ装置内に保持する分散情報が破壊されても、各々のディスク装置ユニットの分散パラメータが正常であれば、ディスク装置ユニットに記録された分散パラメータより、分散情報を再生して、データ破壊を防止する。

【００７４】

【発明の効果】本発明によれば、冗長データのディスク装置への転送時間を短縮したディスクアレイ装置を提供できる。

【００７５】また、障害検出の遅れを防ぎ、信頼性を向上させたディスクアレイ装置を提供できる。

【００７６】また、分散、集合に対する信頼性を向上させたディスクアレイ装置を提供できる。

【図面の簡単な説明】

【図１】冗長データを有するディスクアレイ装置のブロック図である。

【図２】冗長データ生成器のブロック図である。

【図３】ディスクアレイ装置のブロック図である。

【図４】データ分散集合制御回路のブロック図である。

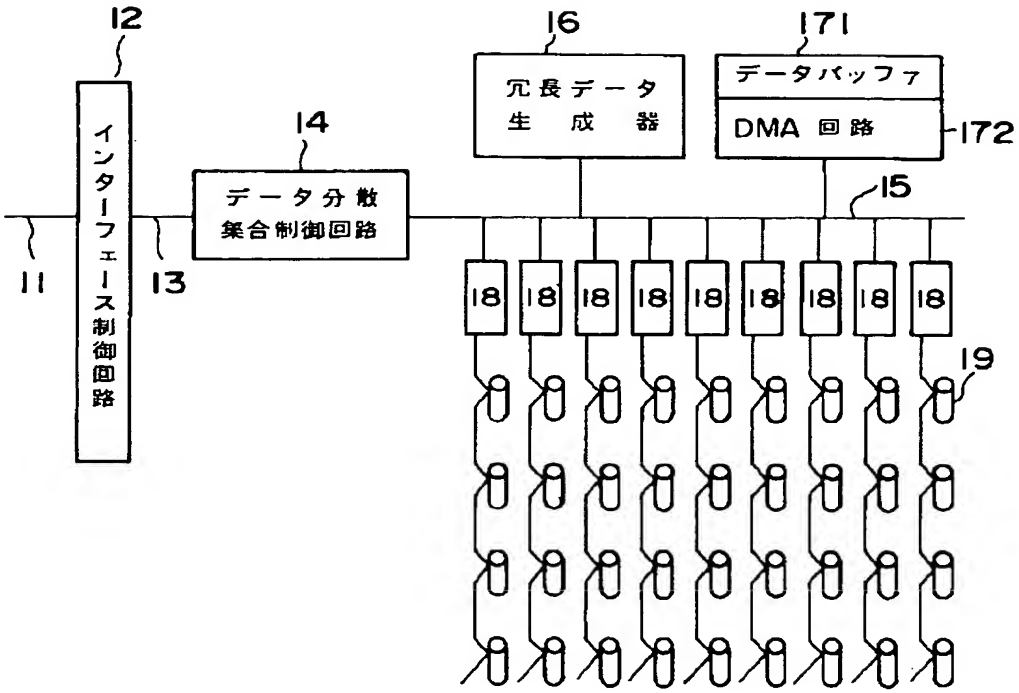
【図５】ディスク制御回路のブロック図である。

【符号の説明】

１１…ホストＳＣＳＩバス、１２…インターフェース制御回路、１３…データバス、１４…データ分散集合制御回路、１５…共通データバス、１６…冗長データ生成器、１７１…データバッファ、１８…ディスク制御回路、１９…ディスク装置ユニット、２１…上位側データ管理カウンタ、２２…ディスク側データ管理カウンタ、２３…冗長データ管理カウンタ、２４…入出力制御回路、２５…入出力可／不可信号。

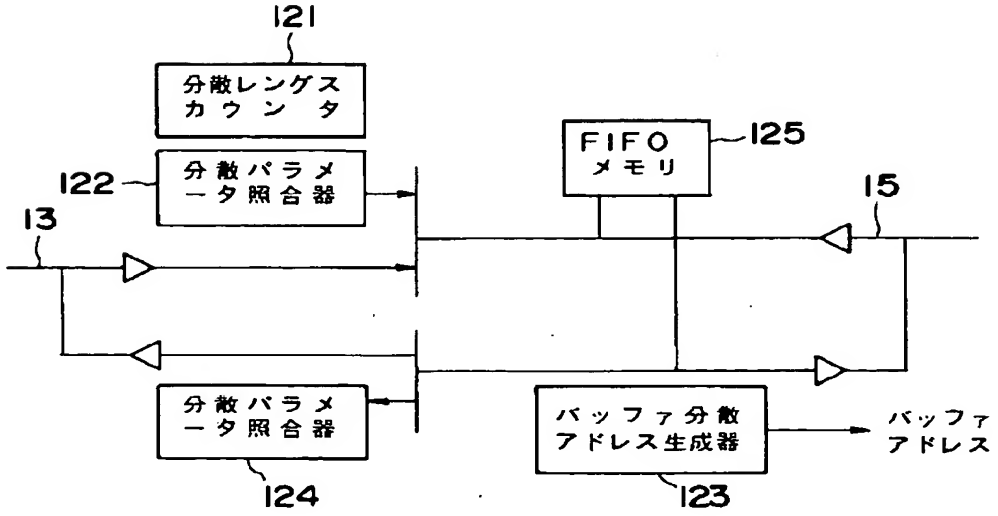
【図１】

ディスクアレイ装置ブロック図（図１）



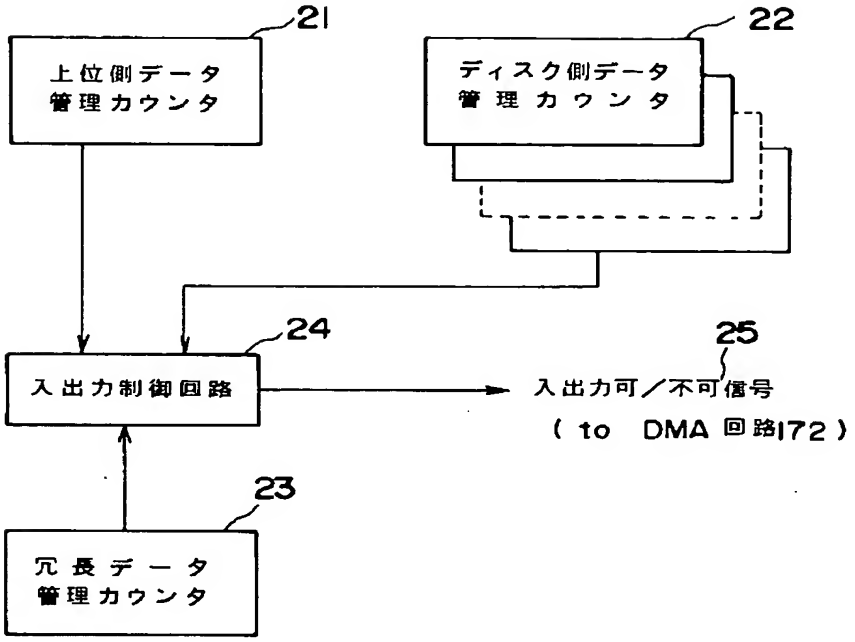
【図４】

図 4



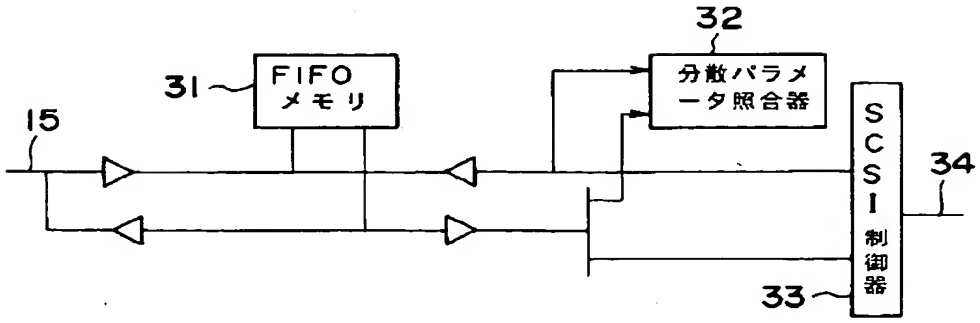
【図２】

冗長データ生成器入出力制御ブロック図（図２）



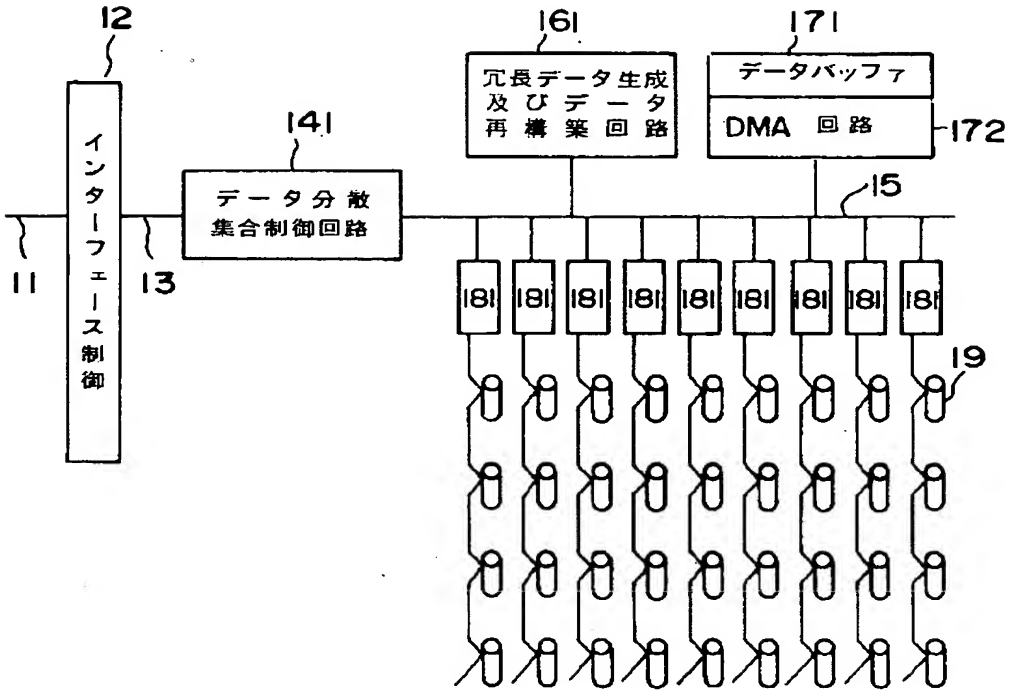
【図５】

図５



【図３】

ディスクアレイ装置ブロック図（図３）



フロントページの続き

(72)発明者　水野　克敏
神奈川県小田原市国府津2880番地　日立コンピュータ機器株式会社内

(72)発明者　鈴木　良一
神奈川県小田原市国府津2880番地　日立コンピュータ機器株式会社内

(72)発明者　馬場　英美
神奈川県小田原市国府津2880番地　日立コンピュータ機器株式会社内